**Year 2022/2023**

COLUMBIA CLIMATE SCHOOL
Climate, Earth, and Society

**INTERNSHIP REPORT**

presented by

**Alexandre REBIÈRE**

From 01/05/2023 to 28/07/2023 at :

**Learning the Earth with Artificial Intelligence & Physics**

LEAP

Internship subject :

# Subseasonal to seasonal climate predictions using AI

Internship Manager : **Pierre GENTINE**

Referring teacher : **Alexandre ALLAUZEN**

---

Columbia University, New York, USA

# Contents


**1 LEAP** **2**

**2 Abstract** **2**

**3 Keywords** **2**

**4 State of the art** **2**

**5 Issue at stake** **2**

**6 Data Management** **2**
6.1 First step : Creating the dataframe . . . 3
6.2 Second step : From dataframe to Pytorch tensor . . . 3

**7 Climate Analysis** **4**
7.1 Analysis of the Western US climate . . . 4
7.2 One interesting climate phenomenon : El Nino . . . 4
7.3 Climate zone separation following Köppen climate classification : . . . 4

**8 Neural Network trials** **5**
8.1 The basic and effective one : One linear layer . . . 6
8.1.1 First approach : grid point by grid point training . . . 6
8.1.2 Second approach : Quicker Training . . . 6
8.2 Introducing non linearity : 2 linear layers and one activation function . . . 6
8.3 Capturing spatial correlations : the use of convolution . . . 7
8.3.1 Trial 1 : Without any reduction of dimension . . . 7
8.3.2 Trial 2 : With reduction of dimension . . . 7
8.3.3 Trial 3 : UNet architecture . . . 7
8.4 Capturing temporal dynamics : LSTM . . . 8

**9 Models performances** **8**

**10 Results** **9**
10.1 Temperature . . . 9
10.1.1 Study of Soil Moisture (CASM) impact . . . 9
10.1.2 Study of elevation impact . . . 11
10.1.3 Study of Gross Primary Production (GPP) and Ecosystem Respiration (RECO) impact . . . 12
10.1.4 SHAP values of the previous features compared . . . 12
10.2 Precipitation . . . 12
10.2.1 Study of Soil Moisture (CASM) impact . . . 13
10.2.2 Study of Gross Primary Production (GPP) and Ecosystem Respiration (RECO) impact . . . 14
10.2.3 Study of elevation impact . . . 14
10.2.4 SHAP values of the previous features compared . . . 14
10.3 Forecasting 3-4 weeks ahead . . . 15

**11 Conclusion, Discussion and Perspectives** **15**

# 1 LEAP

LEAP's primary research strategy is to improve near-term climate projections by merging physical modeling with machine learning across a continuum from expertise in climate science and climate modeling to cutting-edge machine learning algorithms. The benefits will be significant for both the climate and data sciences communities. Climate scientists and modelers struggle to fully integrate the wealth of existing datasets into their models, while machine learning algorithms have been good at emulating and interpolating but have difficulties extrapolating or predicting extremes. By combining both approaches, LEAP will trigger a significant advancement for data science algorithms applied to physical problems. LEAP will incorporate physics and causal mechanisms into machine learning algorithms for better generalization and extrapolation, while optimally using the wealth of data available to climate science, in order to better predict the future.

# 2 Abstract

Weather forecasting has significantly improved in accuracy, thanks to the use of machine learning models and the widespread deployment of weather sensors worldwide. However, predicting sub-seasonal to seasonal climate poses a greater challenge as it no longer relies only on initial conditions. While forecasting the average temperature or precipitation values is relatively easier due to seasonal patterns, predicting temporal and geographical specific anomalies in temperature or precipitation is a different story.

In a context of climate change and the increasing occurrence of droughts, it has become crucial to enhance long-term forecasts of temperature and precipitation. This is essential for adequately preparing for droughts and other extreme weather events. Anticipating these phenomena in advance, along with their intensity, can significantly assist water management efforts and have a positive impact on the economy and the population's well-being.

To enhance the accuracy of long-term forecasts, the U.S. Bureau of Reclamation and the National Oceanic and Atmospheric Administration (NOAA) initiated the Sub-seasonal Climate Forecast Rodeo. This year-long real-time forecasting challenge aims to skillfully predict temperature and precipitation in the western U.S. two to four weeks and four to six weeks in advance *(Hwang, Orenstein, Cohen, Pfeiffer, et al. 2019)*. Here we present and evaluate our machine learning approaches, which are based on the same dataset and data management techniques as the RODEO competition winner. Additionally, we have incorporated new features into our approaches to analyze their impact on forecast accuracy.

Our system consists of two distinct approaches : a machine learning approach utilizing linear regression with XGBoost and a deep-learning approach using various neural networks architectures. Rather than aiming to surpass the state-of-the-art accuracy achieved in the RODEO competition, our primary objective was to develop functional models that allowed us to examine the influence of different new features on forecasts using explainable AI. We aspire that our methodologies contribute to advancing the state of the art in subseasonal forecasting. In this paper, we will only focus on the Deep Learning approach, please look at the work of Alexandre François, the other LEAP intern, to access XGBoost results.

# 3 Keywords

Subseasonal climate forecasting, Temperature, Precipitation, Anomalies, XGBoost, Machine Learning, Deep Learning, Western United States, Drought, Water management, Soil Moisture, Elevation, El Nino, Gross Primary Production, Ecosystem Respiration

# 4 State of the art

In the RODEO competition, forecasts were judged on the spatial cosine similarity between predictions and observations adjusted by a long-term average, the same criteria is used here to evaluate our models' performances. More precisely, let $t$ denote a date represented by the number of days since January 1, 1979, and let *year(t)*, *doy(t)*, and *monthday(t)* respectively denote the year, the day of the year, and the month-day combination (e.g., January 1) associated with that date. We associate with the two-week period beginning on $t$ an observed average temperature or total precipitation $y_t \in R^G$ and an observed anomaly $a_t = y_t - c_{monthday(t)}$
where :

$$c_d \triangleq \frac{1}{30}\Sigma_{t:monthday(t)=d} = y_t, year \in [1981 : 2010]$$

is the *climatology* or long-term average over 1981-2010 for the month-day combination d. Contestant forecasts $\hat{y}_t$ were judged on the cosine similarity-termed skill in meteorology between their forecast anomalies $\hat{a}_t = \hat{y}_t - c_{monthday(t)}$ and the observed anomalies:

$$\boxed{skills(\hat{a}_t, a_t) \triangleq cos(\hat{a}_t, a_t) = \frac{\hat{a}_t, a_t}{\|\hat{a}_t\|_2 \|a_t\|_2}} \quad (1)$$

# 5 Issue at stake

***Can sub-seasonal and seasonal climate predictions be improved by adding new types of data to the model? How do predictions evolve according to the type of data added and their value? What are the advantages and disadvantages of developing a model that prefers Deep Learning to Machine Learning?***

# 6 Data Management

As a foundation for our study, we utilized The Subseasonal Rodeo Dataset *(Hwang, Orenstein, Cohen, and Mackey 2018)*, which was developed by the authors of the paper

we focused on. The objective of our internship was to enhance the model by incorporating new features and studying their impact. Consequently, we had to delve into the data processing techniques employed by the RODEO framework to add these new features to the existing dataframe.

To forecast temperature or precipitation anomalies 5 to 6 weeks in advance, we adopted the method employed by the RODEO candidates, which involved shifting the ground truth data by a specific time period. This required an in-depth analysis of the numerous functions developed by the authors and connecting them together within the architecture to incorporate the new features. We also encountered challenges in temporarily adapting the features since some were measured daily, weekly, monthly, or remained constant over time (e.g. elevation), but found solutions to handle it. Furthermore, some features were directly associated with specific grid points, such as humidity and temperature, while others, like wind, sea surface temperature, and ice concentration, were added as indexes in our database. Consequently, all grid points shared the same value for these features on a given date.

## 6.1 First step : Creating the dataframe

The RODEO models utilized a dataframe object to store data, which was constructed using all the collected feature data. We applied a certain shift to the feature values that we wanted to use for temperature or precipitation forecasting, depending on the prediction we wanted to make. In addition to using the ground truth temperature and precipitation data, we also compared these values to the **"climatology"**, which represents the expected average values for a given date, it reflects what is expected depending of the season you're in. The difference between the ground truth temperature and the climatology data is referred to as **"anomalies"**. In our study, our goal was to forecast temperature or precipitation anomalies, rather than the actual values, 43 days in advance (if predicting at 5-6 weeks) or 29 days in advance (if predicting at 3-4 weeks) . Forecasting temperature values posed fewer challenges due to the predictability associated with seasonal patterns. However, predicting anomalies, which exhibit greater complexity and variability, proved to be more interesting.

To the existing dataframe, we added new features such as Soil Moisture (CASM), Gross Primary Production (GPP), Ecosystem Respiration (RECO), Elevation, and various temporal indexes such as El Nino events and snow cover on other continents, after being pre-processed.

These new dataframes were directly utilized by the models employed in the RODEO paper or the XGBoost model. However, to apply deep learning methods using PyTorch, we had to reshape the data into PyTorch tensors.

## 6.2 Second step : From dataframe to Pytorch tensor

The dataframe contained hundreds of thousands lines of data, with each line corresponding to a specific date and grid point. In total, 514 grid points were considered, represented by integer latitude and longitude values within the Western US zone. To transform the data into a PyTorch tensor, a multidimensional shape was required. We chose the shape (time, latitude, longitude, features) for our tensor. The *"time"* dimension stores all the days considered, from January 1, 1979, to May 23, 2018. The *latitude* and *longitude* dimensions represented the smallest rectangle that enclosed the Western US zone, forming a 23x31 grid, totaling 713 points. However, only 514 of these points were relevant, as the others were located in the Pacific Ocean or outside the Western US borders. During data processing, these irrelevant points were assigned NaN values, which we replaced with zeros. A mask was then applied to exclude these points from model consideration, particularly during loss calculations. Finally, each *feature* was stored in the last dimension of the tensor.

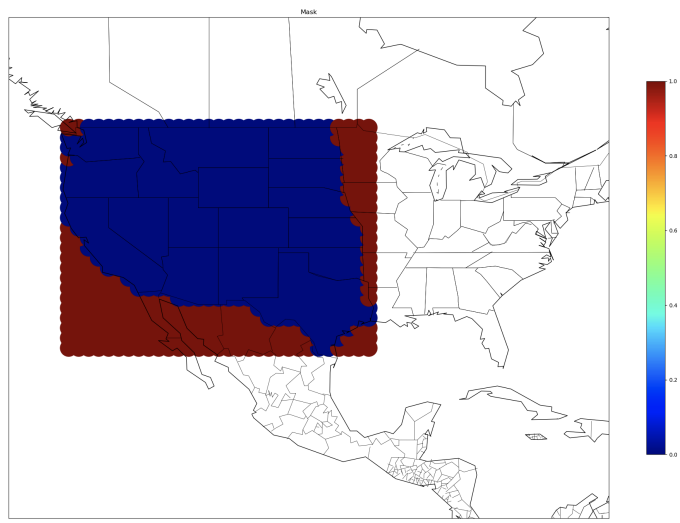


**Figure 1**: Mask applied to the rectangle tensor. In blue are the grid-points taken into account for the loss computation.

Significant efforts were made to handle the addition of features to the tensor. Some features had global coverage with a resolution lower than one degree of latitude and longitude. To address this, we rounded the values to the nearest integer and calculated the mean of the rounded values. In cases where data were missing, we employed a data completion approach. For example, if data was missing for a particular geographical area on a certain date, we used the values from the closest neighbors to estimate the missing value. If all the data for a specific feature were missing on a certain date (e.g., due to weekly or monthly resolution), we copied the previous values to fill in the missing data. Although this approach may introduce some bias, we considered it to be minimal compared to the model's complexity, especially considering the relatively low number of NaN values compared to the overall dataset.

# 7 Climate Analysis

## 7.1 Analysis of the Western US climate

To analyze the outcomes of our models and identify which features had a substantial impact on predictions, we needed a keen interest in climate. The Rodeo competition focused on the Western US zone, which encompassed a diverse range of climates, including dry regions, wet areas, oceanic climates, mountainous terrains, deserts, and agricultural lands. Through our research, we also encountered various climate phenomena, such as El Nino, La Nina, and the Madden-Julian Oscillation (MJO). Also, we observed how certain features beyond the borders of the US could influence precipitation and temperature patterns on American soil.

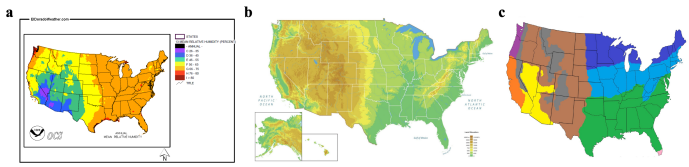


**Figure 2**: **(a)** Mean relative humidity. **(b)** Elevation map of the Western US. **(c)** Climate zones over the US *purple:* marine west coast, *orange:* Mediterranean, *yellow:* mid-latitude desert, *grey:* highland (alpine), *brown:* semi-arid (step), *green:* humid subtropic, *light blue:* humid continental (warm summer), *dark blue:* humid continental (cool summer)

The western US we focus on encompasses some remarkable areas with diverse climates that can have a significant impact on forecasting. Certain features, such as humidity and elevation, can vary drastically from one zone to another. Consequently, their relative weights in the Neural Networks model may differ, as well as their usefulness in making predictions.

The addition of some features might sometimes be useless if they evolve linearly with other features. For example, if soil moisture varies exactly in line with relative humidity in the air, adding soil moisture as a feature would be irrelevant and redundant.

Among the remarkable climate zones over the US we can see a very dry area from east California to west Texas as well as a highland zone over the huge Rocky mountains. Two very wet regions appears in the Western US : the North West near Seattle as well as the East Texas near the Mississippi river. The near Great Lake region is full of crops landscapes with mechanical irrigation in a pretty dry region.

## 7.2 One interesting climate phenomenon : El Nino

The El Niño Southern Oscillation (ENSO) is a cycle of warm and cold phases of sea surface temperature (SST) in the equatorial Pacific Ocean that influences extreme weather and ocean events with significant impacts, such as high temperatures, cold waves, heavy rain, and marine heatwaves (*Kim et al. 2022*).

During normal conditions in the Pacific ocean, trade winds blow west along the equator, taking warm water from South America towards Asia. To replace that warm water, cold water rises from the depths — a process called upwelling. El Niño and La Niña are two opposing climate patterns that break these normal conditions. Scientists call these phenomena the El Niño-Southern Oscillation (ENSO) cycle. El Niño and La Niña can both have global impacts on weather, wildfires, ecosystems, and economies.

Episodes of El Niño and La Niña typically last nine to twelve months, but can sometimes last for years. El Niño and La Niña events occur every two to seven years, on average, but they don't occur on a regular schedule. Generally, El Niño occurs more frequently than La Niña. During El Niño, trade winds weaken. Warm water is pushed back east, toward the west coast of the Americas.

El Niño can affect our weather significantly. The warmer waters cause the Pacific jet stream to move south of its neutral position. With this shift, areas in the northern U.S. and Canada are dryer and warmer than usual. But in the U.S. Gulf Coast and Southeast, these periods are wetter than usual and have increased flooding.

Consequently, ENSO is an important feature to take into account in the dataset we will provide to the model. Besides, this event is so important that the authors already took it into account in their dataset. Nonetheless, with a phenomenon occuring at such a Frequency of 2 to 5 years, only a few example of El Nino events can feed the dataset... Learning will therefore be very limited, and predictions are likely to vary widely, but this phenomenon can weigh heavily on model weights. Phenomena of this type can be the source of instability in the training model.

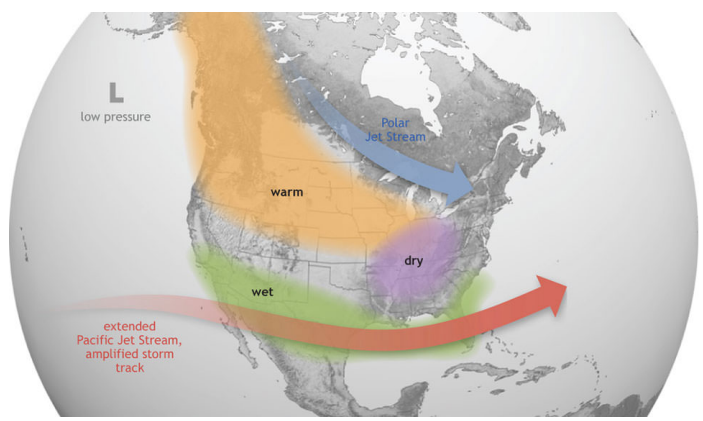


**Figure 3**: El Niño causes the Pacific jet stream to move south and spread further east. During winter, this leads to wetter conditions than usual in the Southern U.S. and warmer and drier conditions in the North. (*National Oceanic and Atmospheric Administration n.d.*)

## 7.3 Climate zone separation following Köppen climate classification :

The Köppen climate classification separates different regions of the World into climate regions where weather conditions are similar. The Western US is then composed of 4 main climates zones (the Polar region isn't representative as it is

composed of only one point over 514). As explained in section 7.1 we observe a very dry region mostly located over the South (Death Valley, deserts in Arizona, Nevada...), Mild temperate climates appears on the North West Coast and near Mississipi, separated between continental and oceanic (wetter) near the Pacific Ocean and the Mexican Gulf. Finally, a continental region is defined over the northern Rocky Mountains and in the North East of the considered zone. Separating these points in climate regions will be useful in the interpretation of the results later. This way, we will be able to analyze geographically the performances of our models, and study the link between forecasting improvements and the added features.

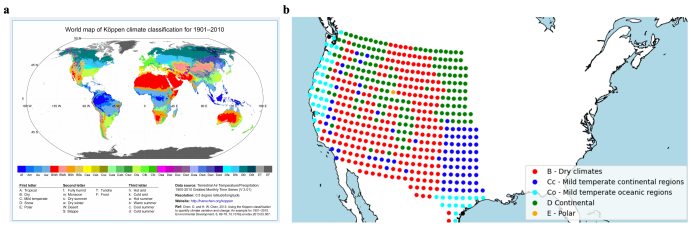


**Figure 4**: **(a)** World map of Köppen climate classification for 1901-2010. **(b)** Simplified western US map of Köppen climate classification for 1901-2010 applied to the grid points studied in this study.

**Personal comment** : *During the internship, my main objective was to utilize simple models to compute new features and assess their impact on forecasting. However, I was also driven by the ambition to achieve superior predictions compared to the state-of-the-art approaches employed in the paper. The absence of Neural Network models in the authors' work, and our quick and successful use of a basic XGBoost algorithm with similar results, further motivated me.*

*To pursue this goal, I devoted about one month of the three-month internship to multiple attempts at implementing a Neural Network algorithm. I wanted this algorithm not only to take into account the features values at the exact same point 43 days ago and perform a linear regression over these features, as it is done in the paper and with XGBoost, but I also wanted it to capture spatial and temporal correlations.*

*I envisioned employing an Auto-Encoder with convolutional layers, adopting a sliding window approach to incorporate feature values from various grid points across the United States. This strategy would enable the model to capture spatial correlations over both short and long distances. Moreover, I sought to extend the temporal scope by incorporating feature values from a certain number of days before D-43. This inclusion of temporal dynamics would prove valuable in forecasting, especially in understanding how temperature anomalies evolve after significant temperature events.*

*To explore these possibilities, I experimented with various architectures, including an LSTM structure, among others. The ultimate goal was to create a more sophisticated and effective model that could surpass the limitations of existing approaches and improve weather forecasting accuracy.*

# 8 Neural Network trials

We attempted several Neural Network architectures before realizing their unproductivity. Instead, we redirected our focus towards studying the impact of features on predictions using simpler models, which yielded favorable results. Our decision to discontinue the research on improving Neural Networks was influenced by reading a paper (He et al. 2021) where the authors emphasized the challenges of Deep Learning models in climate forecasting. Additionally, a discussion with a fellow PhD student, who spent months exploring a Deep Learning approach for her heatwave analysis project before achieving superior results with a basic XGBoost model, further solidified our decision. However, we maintain our confidence that Deep Learning approaches have the potential to deliver even better results with a comprehensive understanding, optimized parameters, and refined datasets.

The listed Neural Network architectures are ranked in order of increasing complexity. With the exception of one linear layer model, none of them yielded interpretable results.

## 8.1 The basic and effective one : One linear layer

**You can find the code of this model in *"linear model predictions.ipynb"*, calling functions from *"tools.py"***

### 8.1.1 First approach : grid point by grid point training

In this approach, our algorithm's objective has been to apply a linear layer to reduce the dimension of the features from *"Number of features"* to *"1"*. To achieve this, a tensor of size (dates, lat, lon, features) (=(14388, 23, 31, features), 14388 corresponds to all the day between 1979 and 2018 available) has been normalized for each features of interest, separated between training, validation and test sets.

These tensors have then been flattened across the spatial dimensions in order to have a shape (dates, lat*lon = 713 , features). However, among the 713 grid points in the rectangle, only 514 are relevant for our study. Hence, we used a linear mask to retain a tensor of size (dates, lat*lon = 514, features).

In this model, the training is date-independent, leading to an input tensor shape of (dates*514, features). The output is a one-dimensional vector (dates*514) containing the predicted temperature or precipitation anomalies. Using this, the output tensor contains a 1D vector of scalar values that able the SHAP (SHapley Additive exPlanations) library to perform explainable AI, useful to understand the role of each feature in the prediction.

However, using this huge vector in this model comes with a significant drawback: computational time. Training the model requires approximately 6 minutes with 5 epochs, and calculating the SHAP values for every grid point and one year of data can take around 30 minutes (about 180,000 SHAP values to compute). As a result, we reserve the use of this model for SHAP values analysis on specific years and features of interest.

For comparing skill improvements when adding new features, we prefer using another model that employs multidimensional tensors as the model's output. This alternative approach strikes a better balance between computational efficiency and the ability to assess the impact of additional features, but appears to have the same skills as the grid point by grid point model.

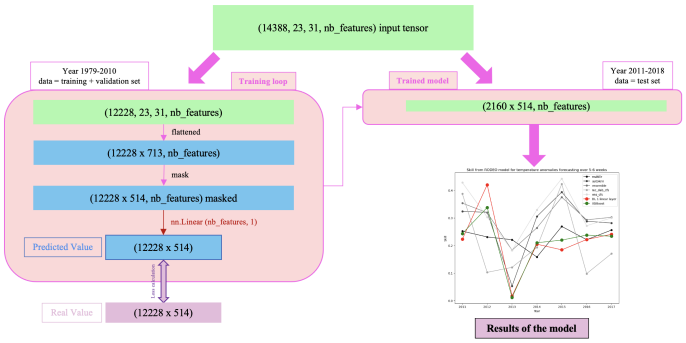


**Figure 5**: Linear model on each grid points

### 8.1.2 Second approach : Quicker Training

The model's approach closely resembles the previous one, with one notable difference: the linear layer now takes a (dates, lat, lon, features) tensor as input and returns a (dates, lat, lon) tensor containing predicted temperatures or precipitation values. The loss calculation is specifically performed only on the relevant grid points of the tensor, thanks to the mask applied.

Unfortunately, the SHAP library cannot be utilized for multidimensional tensors. Instead, we can employ LRP (Layerwise Relevance Propagation) to determine if a feature contributes to the prediction improvement. However, unlike SHAP, LRP does not analyze whether a high value of one feature has a more significant impact on the prediction than a low value.

Despite this limitation, the loss calculation in this model is considerably faster, allowing us to analyze the skills improvement grid point by grid point during the 2011-2017 period. By identifying patterns, we gain a better understanding of why certain features have varying effects on prediction enhancements.

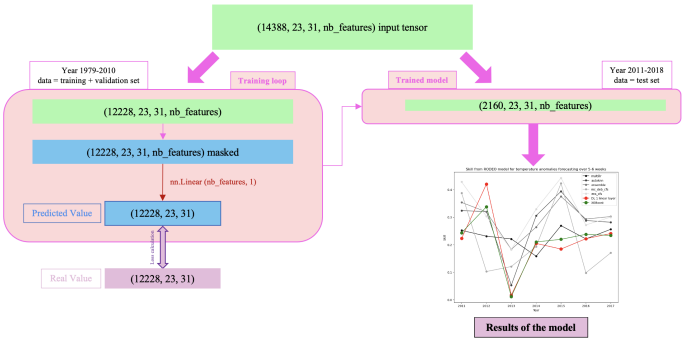


**Figure 6**: Linear model with quicker training

## 8.2 Introducing non linearity : 2 linear layers and one activation function

This architecture is also quite simple. Building upon the second approach mentioned earlier, instead of using a *nn.Linear(features, 1)*, we opted for a sequential approach to incorporate non-linearity: *nn.Sequential( nn.Linear(features, hidden_layer), nn.ReLU(), nn.Linear(hidden_layer, 1))*.

Surprisingly, the resulting skills are notably lower compared to using just one linear layer. Understanding why the introduction of this non-linearity penalizes the predictions to such an extent has proven to be quite challenging.

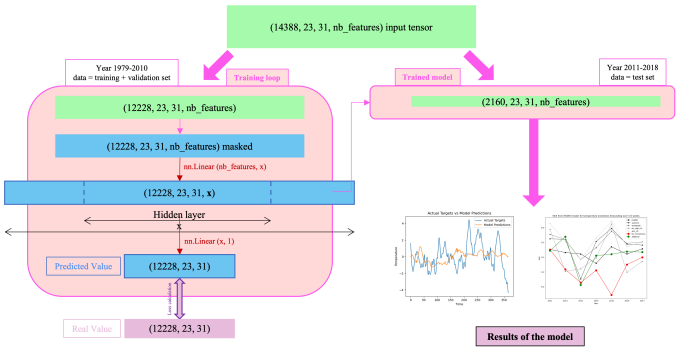


**Figure 7**: Model with non linearity introduced : double linear layer + activation function

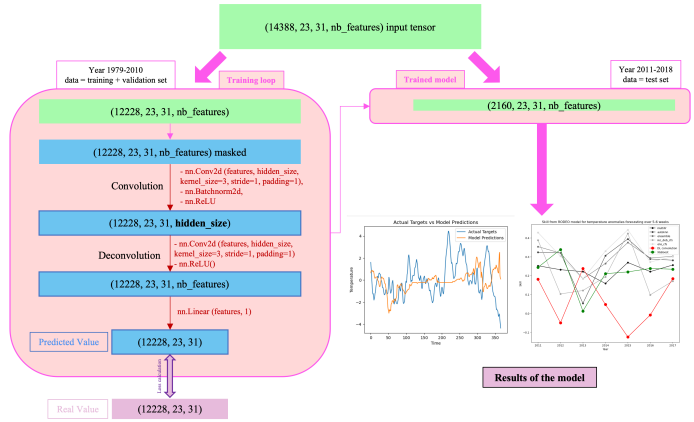


**Figure 8**: Model with simple AutoEncoder to capture spatial correlations

## 8.3 Capturing spatial correlations : the use of convolution

Various approaches have been attempted to capture spatial correlations between grid points with the aim of enhancing predictions. However, the predictions remain constant (when using the same regularization) while the targets vary, leading to poor skills. Unfortunately, none of the methods employed succeeded in altering this phenomenon. Removing the regularization allows predictions to be obtained, but they become highly unstable from one training to another, and the average skill is lower compared to using a simple linear layer. As a consequence, these results are not interpretable.

The primary challenge when applying a sliding window over the grid was precisely defining the border with the mask to ensure that the sliding window treated the US Pacific coast, for example, as an impassable boundary. It was crucial to avoid averaging values on pixels where the feature values were set to zero outside the Western US, as this could introduce biases and unwanted side effects into our predictions.

To address this issue, we employed a solution where we extended (copied) the coastal feature values to the first neighboring pixel outside the Western US. By doing so, we effectively mitigated any side effects that could arise from averaging over values set to zero. Instead of averaging over a zero value, we treated the unused pixel as similar to its closest neighbor, thus preserving the integrity and accuracy of the sliding window computations.

### 8.3.1 Trial 1 : Without any reduction of dimension

The easiest model only uses a sliding window, without reducing dimension

### 8.3.2 Trial 2 : With reduction of dimension

Various techniques, such as *Maxpooling/Upsampling* and varying stride in *Conv2D* or *Conv2Dtranspose*, were employed to reduce and subsequently increase dimensions in an AutoEncoder architecture.

Numerous trials were conducted, incorporating one or more layers of convolution/deconvolution, and the use of batchnorm2d to stabilize calculations. Additionally, different activation functions were tested. We've also tried to apply a linear layer between the convolution and deconvolution layers, we've tried to optimize every parameters, such as batch size, hidden layer, number of layer, disposition of layers, kernel, stride, padding, type of layers, sometimes using Optuna library to optimize it, but neither of these trials had shown great results.

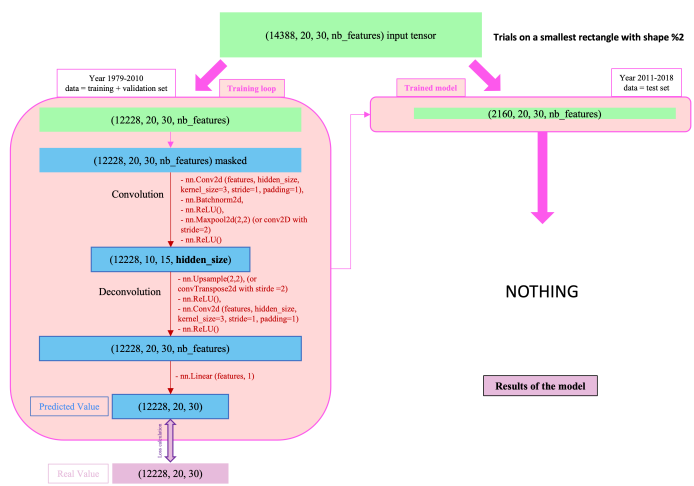


**Figure 9**: Model with deeper AutoEncoder to capture spatial correlations

### 8.3.3 Trial 3 : UNet architecture

Guessing that skipping connections could improve our model results, we've implemented a UNet structure able to skip connections (code below), but again results were not following.

```python
class UNetBlock(nn.Module):
    def __init__(self, in_channels, out_channels):
        super(UNetBlock, self).__init__()
        self.conv1 = nn.Conv2d(in_channels, out_channels, kernel_size=3,
              padding=1)
        self.relu = nn.ReLU(inplace=True)
        self.conv2 = nn.Conv2d(out_channels, out_channels, kernel_size=3,
              padding=1)

    def forward(self, x):
        out = self.conv1(x)
        out = self.relu(out)
        out = self.conv2(out)
        out = self.relu(out)
        return out


class ClimateForecastModel_v3(nn.Module):
    def __init__(self, model_input, hidden_size, hidden_linear, num_layers):
        super(ClimateForecastModel_v3, self).__init__()
        self.downsample_layers = nn.ModuleList()
        self.upsample_layers = nn.ModuleList()

        # Downsample path
        in_channels = model_input
        for _ in range(num_layers):
            out_channels = hidden_size
            self.downsample_layers.append(
            UNetBlock(in_channels, out_channels))
            in_channels = out_channels
            hidden_size *= 2

        # Bottleneck layer
        self.bottleneck = UNetBlock(in_channels, in_channels)

        # Upsample path
        for _ in range(num_layers):
            out_channels = in_channels // 2
            self.upsample_layers.append(
            UNetBlock(in_channels*2, out_channels))
            in_channels = out_channels

        self.final_conv = nn.Conv2d(in_channels, model_input, kernel_size=1)
        self.fc1 = nn.Linear(model_input, hidden_linear)
        self.relu = nn.ReLU()
        self.fc2 = nn.Linear(hidden_linear, 1)

    def forward(self, x):
        skip_connections = []
        x=x.permute(0,3,1,2)
        # Downsample path
        for downsample_layer in self.downsample_layers:
            x = downsample_layer(x)
            skip_connections.append(x)
            x = nn.MaxPool2d(2)(x)
        # Bottleneck layer
        x = self.bottleneck(x)

        # Upsample path
        for upsample_layer, skip_connection in zip(self.upsample_layers,
              reversed(skip_connections)):
            x = nn.functional.interpolate(x, scale_factor=2, mode='nearest')
            x = torch.cat([x, skip_connection], dim=1)
            x = upsample_layer(x)
        x = self.final_conv(x)
        x=x.permute(0,2,3,1)
        # Fully connected layers
        x = self.fc1(x)
        x = self.relu(x)
        x = self.fc2(x)
        x = torch.squeeze(x, dim=3)
        return x
```

Listing 1: UNet skipping connections Trial

### 8.4 Capturing temporal dynamics : LSTM

Another aspect that had not been utilized in the RODEO paper, but has the potential to improve predictions significantly, is capturing temporal correlations and dynamics. Currently, the model only learns to predict 43 days ahead (or 29), using data at day -43 (or -29) from the same grid point. The idea to capture these correlations is to incorporate the feature values not only from day -43 but also from a few days before that (e.g. 5 days). By doing so, the model could potentially grasp temporal dynamics that are valuable for enhancing predictions.

To implement this approach, we introduced an LSTM layer (Long-Short Term Memory), which takes a 3-dimensional tensor as input (batch size, sequence length, features). Here, the sequence length contains the features of the $x$ days we want to study at Day -43 and before. With a sequence length of size $x = 5$, all the feature values from day -43 to day -48 would be stored in this tensor. Since our problem involves spatio-temporal data, the batch size dimension combines dates and all the grid points (datelat-lon), aligning it with the input shape required for an LSTM layer.

Also, we tried to use a 3 dimensional convolution, with a similar method as in section 8.3 but where time is also a screened dimension.

However, employing these approaches led to extensive computation time, even with a very short sequence length. The results obtained were once again unsatisfactory despite the significant computational effort invested.

## 9 Models performances

The primary objective of our work was not to surpass the state-of-the-art performance, but rather to demonstrate the impact of specific features on forecasting. Nonetheless, we needed to assess the accuracy of our models. As a result, we decided to compare our models to the state-of-the-art approaches to gauge whether the skills were similar, considering the years and their corresponding values.

Using the initial features mentioned in the RODEO competition, we managed to obtain skill curves with a similar shape to those of the authors. Although our models were slightly less performant, they remained relevant, as the authors may have dedicated more time to optimizing their models than we did. Overall, the comparison provided valuable insights into the predictive capabilities of our models in relation to the existing state of the art.

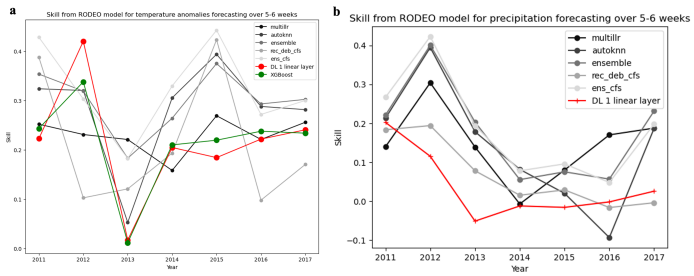


**Figure 10**: Skills distribution of the RODEO models (in gray), XGBoost model (in green) and Deep Learning with one linear layer (in red) for : **(a)** the forecasting of **temperature anomalies** 5-6 weeks ahead, and **(b)** for **precipitation anomalies** 5-6 weeks ahead.

As observed, the skills of both models employed in this article seems to align well with the expected values of skills in the RODEO competition for temperature anomalies forecasting. The skill curve maintains a consistent pattern over the years, indicating a relatively good year in 2012, a notably poor year in 2013, and a series of average years. On the other hand, when it comes to precipitation anomalies, predicting accurately seems to be more challenging. The model developed in this study does not closely match the anticipated skills outlined in the RODEO paper. Despite this, we managed to achieve improved results later on by incorporating new features, leading to substantial enhancements in forecasting performance.

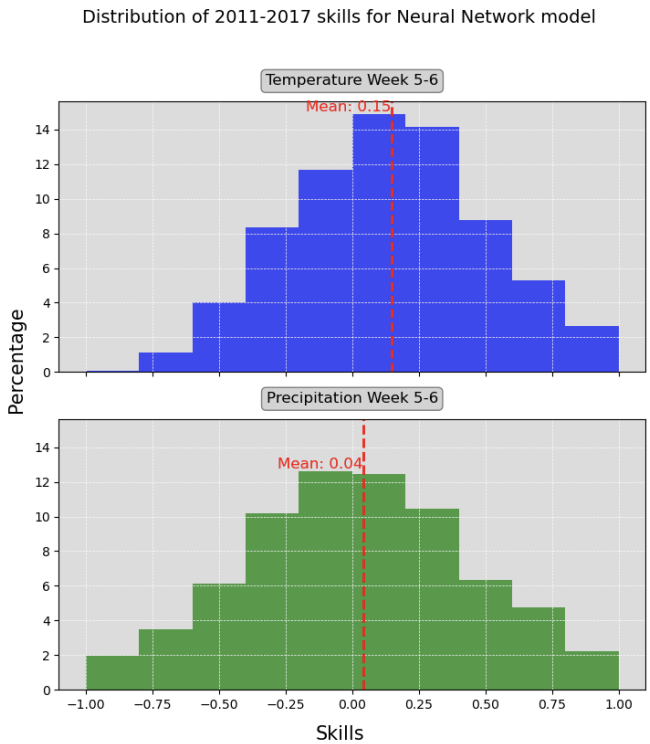


**Figure 11**: Distribution of test set 2011-2017 skills for the NN model composed of one linear layer over the features.

The skills distribution over every grid points at every date predicted by the DL 1 linear layer model shows a better predictability for temperature anomalies than for precipitation anomalies. A behavior that turns out to be rather intuitive in view of the variability of precipitation greater than that of temperatures.

# 10 Results

In this section we evaluate our model forecasts over the Rodeo contest period (2011-2017) and delve into the significant features inferred by each model. Python 3.7 code to reproduce all experiments can be found at https://github.com/AlexandreRebiere/ subseasonal_forecasting_internship. All the plots showed here after were made thanks to the *"Visualization.ipynb"* file calling functions from *"Visualization functions.py"*.

## 10.1 Temperature

**Figures 12 (b), (c), (e), (f)** highlights the low accuracy when predicting anomalies. It is hard to the naked eye to evaluate the performance of the model looking directly at these figures. Only **Figure 12 (c)** seems to show how predictions follow growths or decline trends, even though th prediction curves doesn't really fit the ground-truth values. Skills are then calculated to better assess performances of the model.

We observed that, in average, skills were improved thanks to the addition of new features that seems to be relevant to the predictions (**Figure 12 (a)**). It is very encouraging to see such a room for maneuver in climate predictions accuracy only by adding some well chosen features.

Then, the aim has been to visualize the geographical inequalities in skills improvement and see if it made sense in terms of climate analysis. At first, we tried to observe the apparition of patterns, either on seasons or on climate regions. Then, if an intuition has been made, we decided to quantify this phenomenon. Here after you can find both example of patterns observed, their quantification and their analysis in terms of climate science.

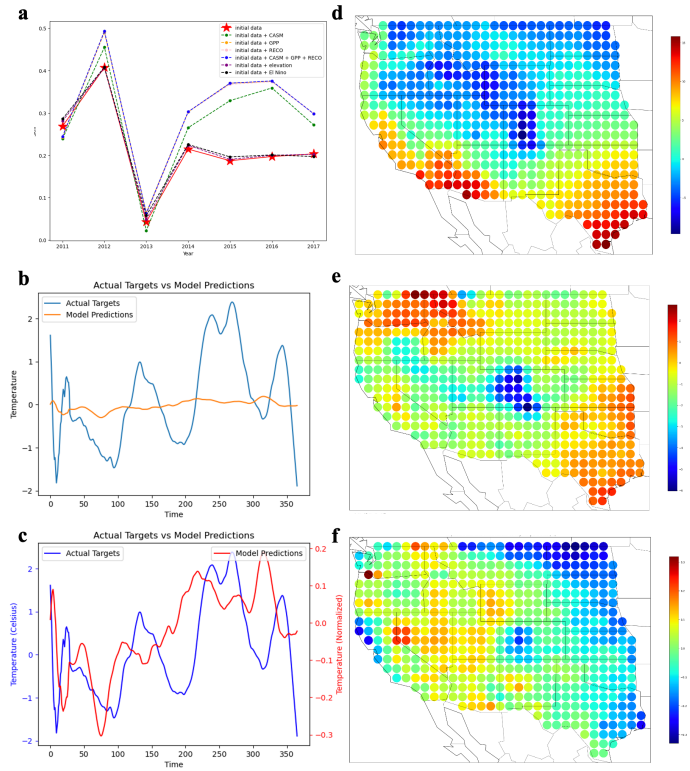


**Figure 12**: Analysis of the model performance after training. All the US maps here have been plotted on a random chosen date : 31st of December 2012, where the value is the average of the 30 previous days to get rid of weather variations. **(a)** Representation of the skills averaged over 5 trainings of the model each time. Each plot represents a certain dataset. The red curve uses the exact same features as the rodeo competition. The other curves are the results of models trained on the same features + new features that had been added. **(b)** Predicted temperature anomalies in 2012 for the grid point of coordinates (lat,lon) = (34,-117). Each point corresponds to the average of the 30 days around the considered date to get rid of small weather variations. **(c)** Same predictions with a zoom on Model Predictions (another scale) **(d)** Temperature on December $31^{st}$ , 2012. **(e)** Temperature anomalies on December $31^{st}$, 2012. **(f)** Predicted temperature anomalies on December $31^{st}$, 2012.

### 10.1.1 Study of Soil Moisture (CASM) impact

In the RODEO competition, only the relative humidity in the air had been taken into account in the features trying to link humidity at day D-43 to the forecast of tempera-

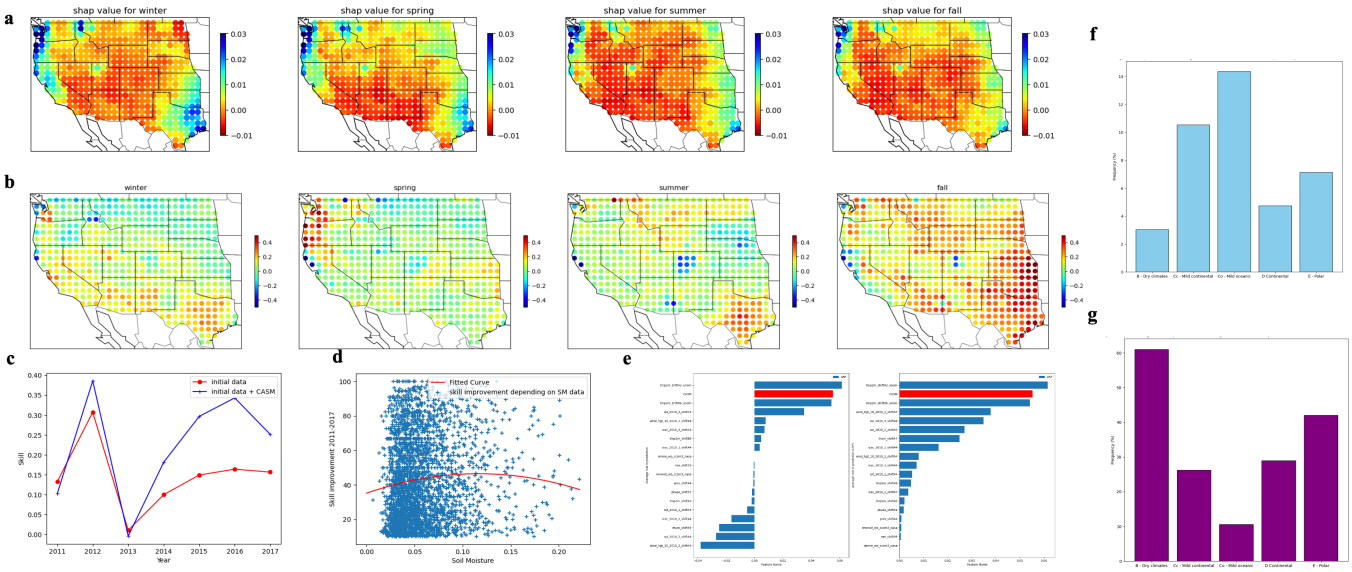


**Figure 13**: Analysis of the impact that had the addition of **soil moisture** to the dataset on **temperature anomalies** predictions 5 to 6 weeks ahead. **(a)** SHAP values for each grid point of the Western US, averaged per season over the year 2016 where the skills improvement had been significant. **(b)** Skill improvement between the model trained on initial data versus the model trained on initial data + Soil Moisture. It has been calculated over each grid point of the Western US, averaged per season over all the years from 2011 to 2017 except 2013 which has been a bad year in prediction accuracy. **(c)** Average skill per year on test set with or without the addition of CASM to the features. **(d)** Representation of the skills improvement regarding the value of soil moisture. Here are only represented the point where there has been a skill improvement between +10% and +100%. On these points it seems that there is an optimum of predictions for a medium range value of soil moisture. But it is difficult to analyze when we consider all the points, it seems instead that predictions increases for high values of soil moisture. This plot might be easily readable for temperature predictions, but seems to be much more chaotic regarding temperature anomalies. **(e)** Layerwise Relevance Propagation values averaged over all the years from 2011 to 2017. A positive value means that the features improves the quality of prediction. *Left :* Real LRP values, *Right :* Absolute LRP values. **(f)** Histogram representing the % of points in climate regions where the skills have been improved by more than 50% with the addition of soil moisture to the dataset. **(g)** Histogram representing % of SHAP values that are lower than 0 per climate region in 2016.

ture or precipitation anomalies. Our intuition has been to think that relative humidity might be much more unstable on long term range than soil moisture. Thus, an anomaly in soil moisture at Day-43 could have a more significant impact on temperature anomalies 43 days ahead. That's why we chose to add a new feature called "CASM" (Consistent AI based Soil Moisture) *Skulovich and Gentine 2022*.

The model has been trained with initial features + soil moisture, then compared to the model trained with initial features. As we can see in **Figure 13 (c)** it improved skills for many years and quite significantly.

According to *Zhang et al. 2020*, due to evaporation and heat fluxes, a higher soil moisture typically causes lower soil and air temperatures, more stable and shallower boundary layers. This leads to a lower cumulus cloud base and higher convective available potential energy. Soil moisture can significantly influence the heat and water exchanges between the atmosphere and the land surface. Therefore, it plays some important roles in numerical weather prediction models. But because here the challenge is to predict temperature anomalies, not temperature, and because predictions are made several weeks in advance, it is not easy to intuit the behavior of soil moisture on our predictions. Do we see higher improvements in wet or dry areas ?

**Figure 13 (e)** and **Figure 16** highlights the role of Soil Moisture in the forecasting made by our new model, as demonstrated by both LRP and SHAP explainable AI methods. Soil moisture proves to have a non-negligible impact on predictions. According to our model predictions, skills were likely to grow when we added "Soil moisture" to the dataset (**Figure 13 (b)**, particularly on oceanic regions and wet areas (mild oceanic and mild continental regions). Indeed, when we analyze the percentage of points per climate areas that have undergone a skill improvement by more than 50% (**Figure 13 (f)** ), it is quite clear that these two climate zones play a vital role in enhancing skill predictions. **Figure 13 (a)** shows the SHAP values over the western US averaged by season. It is an explainable AI method which aims at revealing the weight given to Soil Moisture in predictions, regarding its geographical location. A positive SHAP value shows a correlation between soil moisture and temperature anomalies. It means that in these regions, soil moisture values at D-43 and temperature anomalies at day D are more correlated than elsewhere. Here, it is quite clear that dry areas are anti-correlated to positive anomalies (**Figure 13**

(**a**) and (**g**)). Nonetheless, the higher absolute SHAP values in wet regions shows the model uses soil moisture as an important feature in its predictions in those areas.

One hypothesis that requires confirmation, possibly by employing other Machine Learning models like XGBoost, is that the accuracy of the model may be influenced by the extent of variations induced by a particular feature. For instance, when forecasting temperature anomalies, the model is targeting high-frequency variations, whereas when predicting temperature, the focus is on lower-frequency seasonality. This hypothesis might elucidate why our model assigns greater importance to soil moisture in wet areas. In the following paragraph, you will see that the addition of elevation, a steady feature over time, does not improve significantly the predictions, which seems to corroborate this hypothesis.

### 10.1.2 Study of elevation impact

To the initial data we decided to add information about the altitude of the geographical points we were studying. This feature remains invariant through time and serves as an additional piece of information that could potentially enhance our model's predictive capabilities. Our intuition was that this additional data would significantly improve temperature predictions since temperature is indeed influenced by altitude.

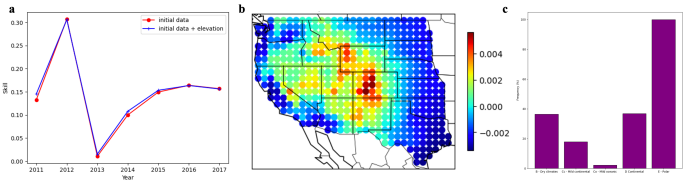


**Figure 14**: Analysis of the impact that had the addition of elevation to the dataset on temperature anomalies predictions 5 to 6 weeks ahead. **(a)** Skill comparison between the model trained on initial data versus the model trained on initial data + elevation. It has been calculated over each grid point of the Western US, averaged per season over all the years from 2011 to 2017 except 2013 which has been a bad year in prediction accuracy. **(b)** SHAP values for each grid point of the Western US, averaged over the year 2016. **(c)** Histogram representing % of SHAP values that are higher than 0.001 per climate region in 2016.

What we weren't sure about, however, was whether this new data could impact our predictions on temperature anomalies. And visibly, we had reasons to doubt, because as we examined the model's average skills over the 2011-2017 period, it became apparent that the addition of elevation

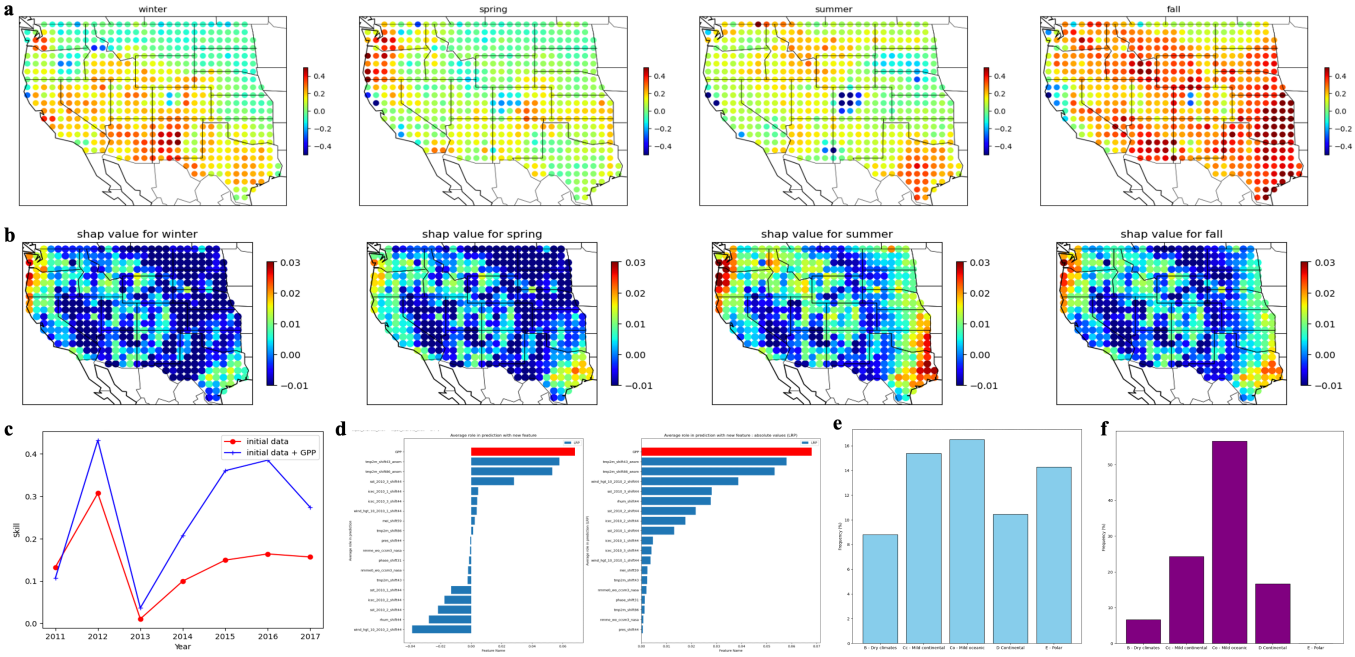


**Figure 15**: Analysis of the impact that had the addition of **GPP (Gross Primary Production)** to the dataset on **temperature anomalies** predictions 5 to 6 weeks ahead. **(a)** Skill improvement between the model trained on initial data versus the model trained on initial data + GPP. It has been calculated over each grid point of the Western US, averaged per season over all the years from 2011 to 2017 except 2013 which has been a bad year in prediction accuracy. **(b)** SHAP values for each grid point of the Western US, averaged per season over the year 2016 where the skills improvement had been significant. **(c)** Average skill per year on test set with or without the addition of GPP to the features. **(d)** Layerwise Relevance Propagation values averaged over all the years from 2011 to 2017. A positive value means that the features improves the quality of prediction. *Left :* Real LRP values, *Right :* Absolute LRP values. **(e)** Histogram representing the % of points in climate regions where the skills have been improved by more than 50% with the addition of GPP to the dataset. **(f)** Histogram representing % of SHAP values that are higher than 0.01 per climate region in 2016.

data to the dataset only resulted in negligible improvements in skills. It appears that including a constant feature does not assist the model in predicting variations with high temporal frequency and amplitude.

Nevertheless, what is interesting to notice is that the SHAP values, while they do not contribute significantly to improving the model's predictions, are higher in the Rocky Mountains region (**Figure 14 (b)**) compared to other areas, as also confirmed in **Figure 14 (c)**. This suggests that a higher altitude encourages positive anomalies in temperature, even though the correlation remains very low.

#### 10.1.3 Study of Gross Primary Production (GPP) and Ecosystem Respiration (RECO) impact

Both features were part of a same dataset called "Metaflux" *Nathaniel, Liu, and Gentine 2023* we decided to add to the initial data used in RODEO to study their effects on predictions. GPP (for Gross Primary Production) corresponds to the total amount of carbon compounds produced by photosynthesis of plants in an ecosystem in a given period of time. Ecosystem respiration (RECO) is the process by which all organisms in an ecosystem convert organic carbon to CO2. As observed in **Figure 12 (a)**, both GPP and RECO significantly improved the skills predictions. It appears that these features have a similar impact on the improvement of skills, both geographically and temporarily, suggesting a potential high correlation between them. In **Figure 15** are thus only represented skills improvement analysis for GPP while we have similar results for RECO.

Because Gross Primary Production and Ecosystem Respiration are high in regions with high vegetation, we had the intuition that it would be in the wet regions (mild oceanic mostly but also mild continental regions), where vegetation is more dense, that the impact on predictions would be higher. This intuition appears to be partially verified, as depicted in **Figure 15 (e)**, where a higher proportion of points experienced a skills improvement of more than 50% in the mentioned regions. This assumption could be discussed depending on the seasons. Indeed, when averaged over the years 2011-2017, it seems that the region where skills improved the most in winter is located in the dry South (**Figure 15 (a)** ). When examining the SHAP values, it is very clear that the highest correlation between GPP and temperature anomalies is located in mild oceanic regions (**Figure 15 (b)** and **(f)** ).

#### 10.1.4 SHAP values of the previous features compared

This **Figure 16** demonstrates the leading roles of the addition of CASM, GPP, RECO and the poor role of elevation in temperature anomalies prediction. The indexed features (such as wind, sea surface temperature, El Nino etc) have voluntarily been removed from the plot to gain clarity. A positive SHAP value indicates that an increase in the feature is correlated with the increase of the targeted feature (here temperature anomalies). Here the absolute value is interesting, a high absolute value indicates a leading role in the model prediction, it doesn't mean the skills will be enhanced significantly, but are likely to.

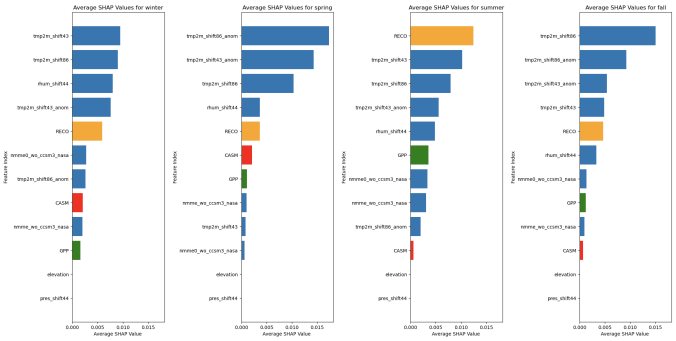


**Figure 16**: Absolute values of SHAP values over each season averaged for all the years from 2011 to 2017.

## 10.2 Precipitation

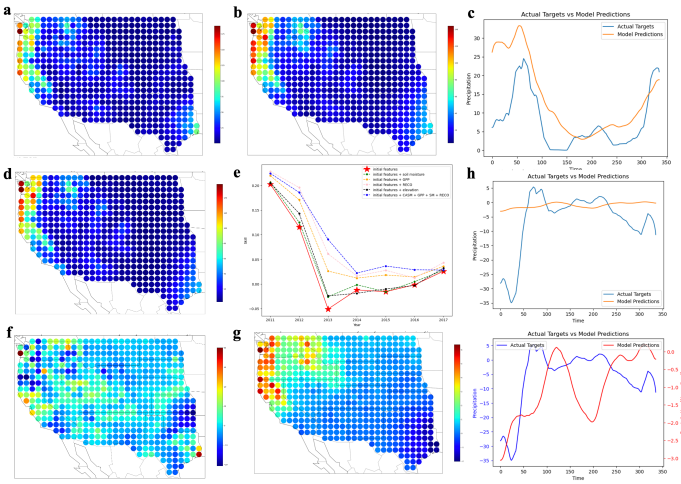


**Figure 17**: Analysis of the model performance after training. Comparison between precipitation predictions and precipitation anomalies predictions. All the US maps and plots here have been plotted on a random chosen date : $31^{st}$ of December 2012, where the value is the average of the 30 previous days to get rid of weather variations. **(a)** Real precipitation. **(b)** Precipitation predicted. **(c)** Precipitation predictions for the grid point of coordinates (lat,lon) = (34,-117) over all the year 2012. **(d)** Precipitation climatology (what should be expected following seasonality) **(e)** Representation of the skills averaged over 5 trainings of the model each time. Each plot represents a certain dataset. The red curve uses the exact same features as the rodeo competition. The other curves are the results of models trained on the same features + new features that had been add. **(f)** Real precipitation anomalies. **(g)** Predicted precipitation *anomalies*. **(h)** Precipitation *anomalies* predictions for the grid point of coordinates (lat,lon) = (34,-117) over all the year 2012, (top) predictions in orange, (bottom) same predictions in red but on another scale.

Precipitation anomalies we want to predict (**Figure 17 (f)**) is the result of the difference between precipitation values (**Figure 17 (a)**) and precipitation climatology (**Figure**

**17 (d)**), it means the difference between the averagely expected value of precipitation and what really happened. When training the model to predict precipitation 43 days in advance, we got quite excellent results (**Figure 17 (b)** and **(c)**). Indeed, even though predictions are not perfect, the amount of precipitation over the year follows a certain pattern that looks like seasonality, and the amplitude of prediction is satisfying, we don't have to use another scale to see them. Using precipitation as a target, we get an average skill value over the years at 0.75 which is very high, even more when we compare it to the skills obtained when using this time precipitation *anomalies* as a target of the model (**Figure 17 (e)**) where averaged skills struggles to surpass 0.05.

However, the predictions for precipitation anomalies (Figure **17 (g)** and **(h)**) appear to be challenging. The model faces more difficulty in accurately predicting precipitation anomalies compared to temperature anomalies, a behavior also observed by the winners of the RODEO competition in their model. The amplitude of predictions for precipitation anomalies is much lower than the target values, making it difficult to achieve precision. Precipitation anomalies eliminate any seasonality aspect, and geographically speaking, the predictions lack precision. Consequently, the model's skills deteriorate significantly. Therefore, our focus in this work was not solely on improving the model to achieve higher skills. Instead, we kept this model and explored the addition of different features to determine their relevance in improving our results, using explainable AI as an additional tool.

**Figure 17 (e)** clearly shows that, in average, skills are improved thanks to the addition of new features that seems to be relevant to the predictions. Then, the aim has been to visualize the geographical inequalities in skills improvement and to see if it makes sense in terms of climate analysis. At first, we tried to observe the apparition of patterns, either on seasons or on climate regions. Then, if an intuition has been made, we decided to quantify this phenomenon. Here after you can find both example of patterns observed, their quantification and their analysis in terms of climate science.

### 10.2.1 Study of Soil Moisture (CASM) impact

Similarly to what has been done for temperature anomalies forecasting, here are presented the outcomes of the addition of soil moisture to our dataset and its impact on forecast.

The addition of Soil Moisture in the dataset had a positive impact on skills predictions in average. The elements visible in **Figure 18 (a)** and **(e)** seems to show an improvement in predictions mostly in wetter regions such as mild oceanic and continental regions. Looking at the SHAP val-

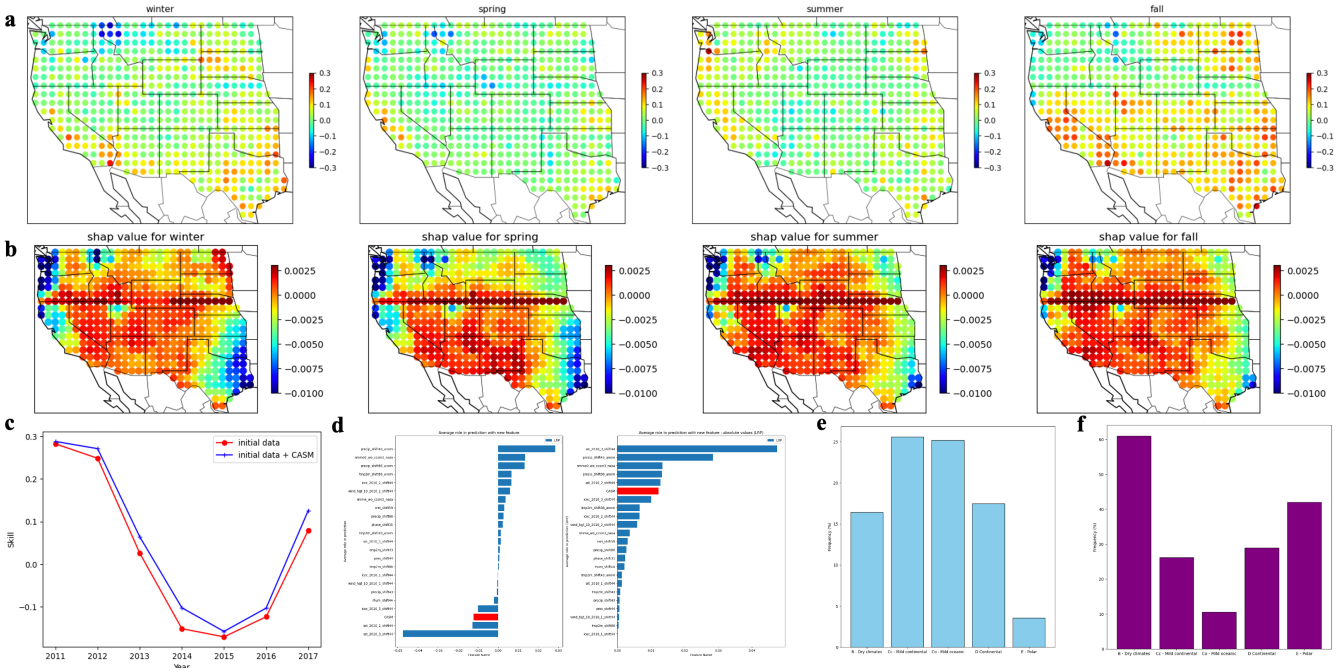


**Figure 18**: Analysis of the impact that had the addition of **soil moisture** to the dataset on **precipitation anomalies** predictions 5 to 6 weeks ahead. **(a)** Skill improvement between the model trained on initial data versus the model trained on initial data + Soil Moisture. It has been calculated over each grid point of the Western US, averaged per season over all the years from 2011 to 2017. **(b)** SHAP values for each grid point of the Western US, averaged per season over the year 2012. **(c)** Average skill per year on test set with or without the addition of CASM to the features. **(d)** Layerwise Relevance Propagation values averaged over all the years from 2011 to 2017. A positive value means that the features improves the quality of prediction. *Left :* Real LRP values, *Right :* Absolute LRP values. **(e)** Histogram representing the % of points in climate regions where the skills have been improved by more than 10% with the addition of soil moisture to the dataset. **(f)** Histogram representing % of SHAP values that are higher than 0 per climate region in 2016.

ues for CASM in the model prediction of prediction anomalies, these SHAP values appears negative, which means that a high soil moisture value is likely to be correlated with a negative anomaly in precipitation in those wet regions. Nonetheless, regarding the absolute values of SHAP for soil moisture, it is clear that the model uses more this feature in wet regions such as mild oceanic and continental regions (**Figure 18 (b)** and **(f)**).

### 10.2.2 Study of Gross Primary Production (GPP) and Ecosystem Respiration (RECO) impact

Similarly to what has been done before in Temperature predictions, both GPP and RECO features have similar behaviors. In **Figure 19** are thus only represented skills improvement analysis for RECO while we have similar results for GPP.

### 10.2.3 Study of elevation impact

Similarly to what has been found in temperature predictions, elevation plays a negligible role in predictions, skills are not improved significantly **Figure 20**. Nonetheless, the model gives more weight to this feature in the Rocky mountains, where elevation is high. Thus, a correlation, even extremely low, had been found by the model between elevation and precipitation anomalies 43 days ahead.

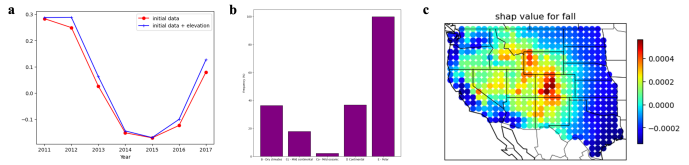


**Figure 20**: Analysis of the impact that had the addition of elevation to the dataset on precipitation anomalies predictions 5 to 6 weeks ahead. **(a)** Skill comparison between the model trained on initial data versus the model trained on initial data + elevation. It has been calculated over each grid point of the Western US, averaged per season over all the years from 2011 to 2017 . **(b)** Histogram representing % of SHAP values that are higher than 0.0001 per climate region in 2012. **(c)** SHAP values for each grid point of the Western US, averaged over the year 2012.

### 10.2.4 SHAP values of the previous features compared

This Figure demonstrates the leading roles of the addition of CASM, GPP, RECO and the poor role of elevation in precipitation anomalies prediction. Again, the indexed features (such as wind, sea surface temperature, El Nino etc) have voluntarily been removed from the plot to gain clarity.

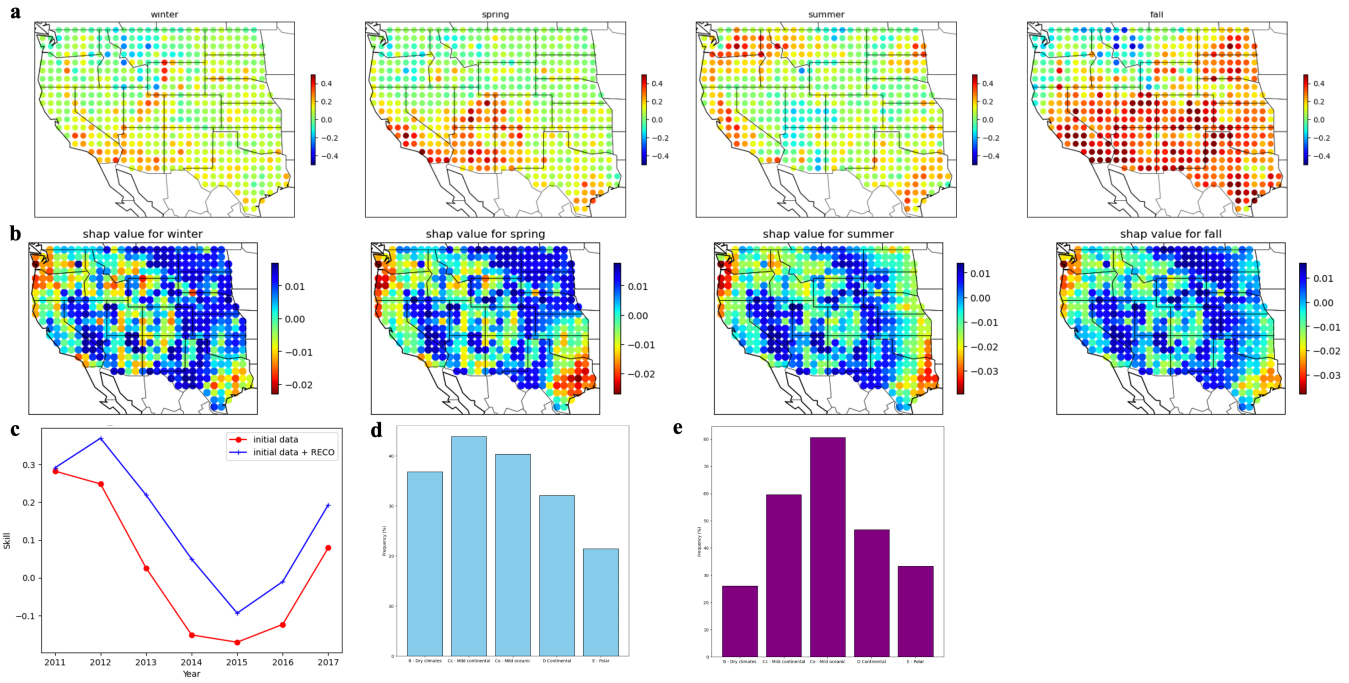


**Figure 19**: Analysis of the impact that had the addition of **RECO (Ecosystem Respiration)** to the dataset on **precipitation anomalies** predictions 5 to 6 weeks ahead. **(a)** Skill improvement between the model trained on initial data versus the model trained on initial data + RECO. It has been calculated over each grid point of the Western US, averaged per season over all the years from 2011 to 2017. **(b)** SHAP values for each grid point of the Western US, averaged per season over the year 2012 where the skills improvement had been significant. **(c)** Average skill per year on test set with or without the addition of RECO to the features. **(d)** Histogram representing the % of points in climate regions where the skills have been improved by more than 10% with the addition of RECO to the dataset. **(e)** Histogram representing % of SHAP values that are lower than -0.002 per climate region in 2012.

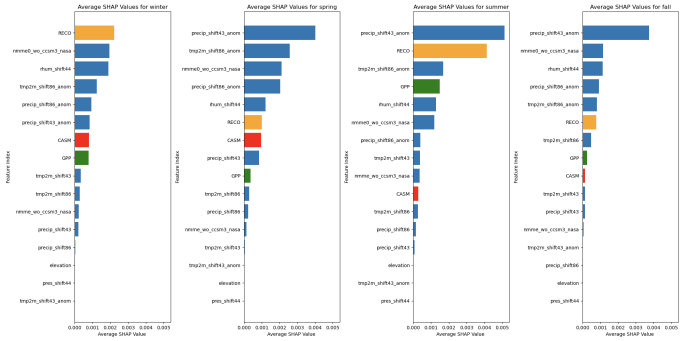


**Figure 21**: SHAP values showing the leading roles of the addition of CASM, GPP, RECO

### 10.3 Forecasting 3-4 weeks ahead

Similar plots have been plotted when predicting precipitation and temperatures anomalies 3-4 weeks in advance, instead of 5-6 weeks. To do so, data has been shifted by 29 days instead of 43 days in the dataset. Similar results have been observed, the behavior of the models towards the features doesn't change when forecasting at a closer date, even if skills are significantly improved (**Figure 22**), with the same shape over the years.

| Prediction \ Target Horizon | 5-6 weeks | 3-4 weeks |
|---|---|---|
| Temperature anomalies | 0.2162 | 0.2461 (+13.8%) |
| Precipitation anomalies | 0.0377 | 0.0575 (+52.5%) |

**Figure 22**: Skill average value over 2011-2017.

## 11 Conclusion, Discussion and Perspectives

In this study, we have demonstrated the potential for improving climate predictions by strategically incorporating new data into our models. Starting with a baseline model consisting of a one linear layer Neural Network, the addition of each carefully chosen feature to the initial dataset resulted in varying degrees of performance improvement. This reveals a wide margin of maneuver for enhancing climate predictions by judiciously selecting and augmenting data, thereby hoping to approach the levels of improvement seen in weather forecasts in recent decades.

Furthermore, we established that the enhancement of skills was dependent on both time and spatial factors. The skills improvements upon adding features were not uniform across the Western US territory, as the model assigned varying weights to new features in different regions, impacting its predictions accordingly. Our work also focused on interpreting these phenomena through a climate analysis approach to validate or challenge our initial assumptions. To this end, we analyzed the US western territory using climate zones based on the Köppen climate classification. While temperature and precipitation predictions exhibited comprehensible behaviors, anomalies proved to be less intuitive. Some behaviors were likely to be different between the Deep Learning one linear layer used here and the XGBoost model used by my fellow intern. For instance, the XGBoost model placed a greater emphasis on elevation for skills enhancement, while GPP and RECO played more significant roles in the Deep Learning model. Additionally, spatial behaviors between the two models appeared to differ. These findings may motivate future researchers in LEAP to explore combining these models to further enhance predictions, leveraging the strengths of each model for higher accuracy.

Finally, we remain optimistic about the potential of Deep Learning in future climate predictions. While our trials revealed the high variability of Deep Learning predictions, we believe that with larger datasets, better optimization, finer resolutions, and broader study areas encompassing the entire globe, Deep Learning algorithms can capture spatial and temporal correlations using techniques such as convolution and LSTM layers that couldn't be captured using a basic Machine Learning linear regression. Because unlike to weather predictions, climate predictions are made on long time range and require consideration of conditions at a larger geographical scale, an ambitious project would involve harnessing the power of Transformers to capture long-distance spatial correlations using data from all around the Earth, leading to even more advanced and accurate climate predictions.

## References

[He+21] Sijie He et al. "Sub-seasonal climate forecasting via machine learning: Challenges, analysis, and advances". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 1. 2021, pp. 169–177.

[Hwa+18] Jessica Hwang, Paulo Orenstein, Judah Cohen, and Lester Mackey. "The SubseasonalRodeo Dataset". Version V2. In: (2018). DOI: 10.7910/DVN/IHBANG. URL: https://doi.org/10.7910/DVN/IHBANG.

[Hwa+19] Jessica Hwang, Paulo Orenstein, Judah Cohen, Karl Pfeiffer, et al. "Improving Subseasonal Forecasting in the Western U.S. with Machine Learning". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD '19. Anchorage, AK, USA: Association for Computing Machinery, 2019, pp. 2325–2335. ISBN: 9781450362016. DOI: 10.1145/3292500.3330674. URL: https://doi.org/10.1145/3292500.3330674.

[Kim+22] Jinah Kim et al. "Spatiotemporal neural network with attention mechanism for El Niño forecasts". In: *Scientific Reports* 12.1 (2022), p. 7204.

[NLG23] Juan Nathaniel, Jiangong Liu, and Pierre Gentine. "MetaFlux: Meta-learning global carbon fluxes from sparse spatiotemporal observations". Version 1. In: (Mar. 2023). DOI: 10.5281/zenodo.7761881. URL: https://doi.org/10.5281/zenodo.7761881.

[Nat] National Oceanic and Atmospheric Administration. *El Niño and La Niña*. https://oceanservice.noaa.gov/facts/ninonina.html. Accessed on July 24, 2023.

[SG22] Olya Skulovich and Pierre Gentine. "CASM: A Long-term Consistent Artificial Intelligence-based Soil Moisture Dataset". In: *AGU Fall Meeting Abstracts*. Vol. 2022. 2022, H25R–1316.

[Zha+20] Hailiang Zhang et al. "The Impacts of Soil Moisture Initialization on the Forecasts of Weather Research and Forecasting Model: A Case Study in Xinjiang, China". In: *Water* 12.7 (2020). ISSN: 2073-4441. DOI: 10.3390/w12071892. URL: https://www.mdpi.com/2073-4441/12/7/1892.